# ABSTRACT

A system and method facilitating incremental web crawl(s) using chunk(s) is provided. The system can be employed, for example, to facilitate a web-crawling system that crawls (*e.g.*, continuously) the Internet for information (*e.g.*, data) and indexes the information so that it can be used as part of a web search engine.

The system facilitates incremental re-crawls and/or selective updating of information (*e.g.*, documents) using a structure called a chunk to simplify the process of an incremental crawl. A chunk is a set of documents that can be manipulated as a set (*e.g.*, of up to 65,536 (64K) documents). "Document" refers to a corpus of data that is stored at a particular URL (*e.g.*, HTML, PDF, PS, PPT, XLS, and/or DOC Files etc.)

A chunk is created by an indexer. The indexer can place into a chunk documents that have similar property(ies). These property(ies) include but are not limited to: average time between change and average importance. These property(ies) can be stored at the chunk level in a chunk map. The chunk map can then be employed (*e.g.*, on a daily basis) to determine which chunk(s) should be re-crawled.